The Broad Street Pump

March 2023

Issue 58

A Centre for Infectious Diseases and Microbiology - Public Health (CIDM-PH) and Sydney Institute for Infectious Diseases (Sydney ID) publication

Guiding the design of SARS-CoV-2 genomic surveillance by estimating the resolution of outbreak detection

Carl J.E. Suster [1,2], Alicia Arnott [1,2,3], Grace Blackwell [2,3], Mailie Gall [2,3], Jenny Draper[2,3], Elena Martinez[2,3], Alexander P. Drew[2,3], Rebecca J. Rockett[1,2], Sharon C.-A. Chen [1,2,3], Jen Kok [1,2,3], Dominic E. Dwyer [1,2,3] and Vitali Sintchenko [1,2,3]

[1] *Centre for Infectious Diseases and Microbiology–Public Health, Westmead Hospital, Westmead, New South Wales, Australia*

[2] *Sydney Infectious Diseases Institute, The University of Sydney, Camperdown, New South Wales, Australia*

[3] *Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Westmead, New South Wales, Australia*

Published article: https://doi.org/10.3389/fpubh.2022.1004201

The COVID-19 pandemic has demonstrated more clearly than ever before the utility of integrated genomic surveillance in mounting an effective public health response. With unprecedented numbers of genomes it has been possible to track the evolution of SARS-CoV-2 as new variants emerge, and to fill in gaps in conventional epidemiological data to resolve disease transmission chains and outbreak clusters. In New South Wales, the early waves of SARS-CoV-2 in 2020 and 2021 were characterised by limited community transmission and by detailed epidemiological data facilitating high resolution reconstruction of outbreak clusters[1]. Integrated genomic surveillance supported outbreak investigations by linking cases with incomplete epidemiological data to known clusters, monitoring the evolution of local clusters, and corroborating evidence for specific transmission pathways[2–4].

Inside this issue

Dr Carl Suster

Postdoctoral Research Fellow

CIDM-PH

The arrival in NSW of the Delta variant followed by the Omicron variant in late 2021 saw the first widespread community transmission. The high surveillance resolution of the early pandemic became unsustainable as the escalating caseload forced public health units and laboratories to target their resources towards a subset of reported cases. With limited ability to reconstruct outbreak clusters, public health priorities turned to monitoring the emergence and propagation of viral variants in the community using representative genomic surveillance. In tandem, targeted whole genome sequencing (WGS) was employed to support urgent investigations such as outbreaks in high-risk settings including aged care facilities or potential hotel quarantine breaches.

The question of how to direct limited WGS resources for maximum public health impact has become increasingly urgent internationally as the pandemic progresses. Guidelines published by the European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO) acknowledged the need for a combination of targeted and representative WGS[5,6]. They included a method for estimating the number of samples that should be sequenced in a given time period based on the total community incidence and desired precision to which the proportion of viral variants circulating should be known. While this method is a useful starting point, it is based on an assumption of representative sampling from community cases which is rarely achievable in realistic settings due to factors including logistical constraints and inequities in access to diagnostics. Furthermore, the method does not support planning for other public health objectives including cluster resolution.

In this study, we outline an approach to guiding the design of genomic surveillance systems that overcomes some of these limitations. Our method estimates the likelihood of resolving clusters for different scenarios, and explores the impact of alternative sampling strategies. This allows for quantitative estimates of the depth of genomic surveillance required to maintain the high-resolution reconstruction of transmission clusters, and an explicit means to model certain sampling biases. The method works by using an algorithm to cluster cases based on their genetic and temporal similarity, and then comparing these reconstructed clusters to the composition of known clusters. Where the two cluster definitions are concordant, we infer that high resolution clustering is viable.

Simulated outbreaks

The surest way to establish a ground truth for cluster definitions is to simulate outbreaks, enabling perfect access to the "true" transmission events linking cases. We used a simple branching process to generate plausible transmission trees for SARS-CoV-2 using crudely realistic parameters for the temporal progression of each infection and the reproduction number. We then overlaid a simple representation of the viral genome onto these trees and used knowledge of the infector of each case and time between infections to randomly introduce point mutations at the appropriate rate. Finally, we applied a small random reporting delay. The simulation yields the pairwise temporal distance in reporting dates, and the pairwise genetic distance in terms of the number of point mutations (Figure 1A-B).

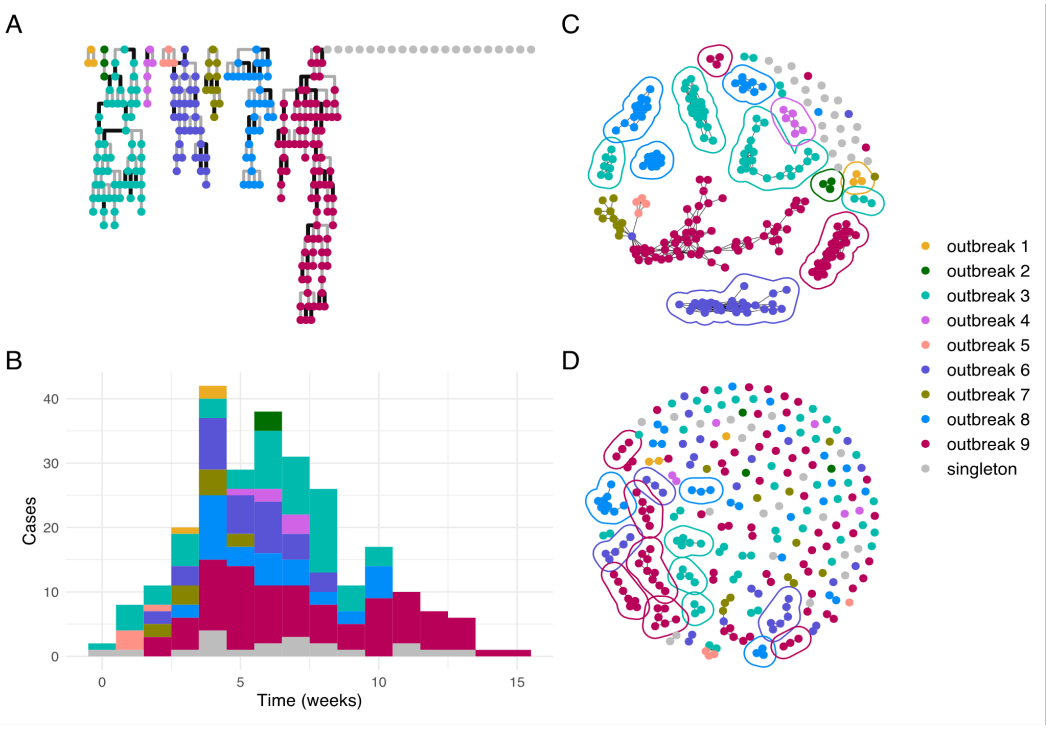

Figure 1: An example simulation showing (A) the generated transmission trees and (B) the epidemiological curve. Distinct outbreaks are coloured consistently between panels. The reconstructed clusters are shown by groups of cases connected by lines with informative clusters enclosed in a coloured outline for (C) the threshold method and (D) the kNN method.

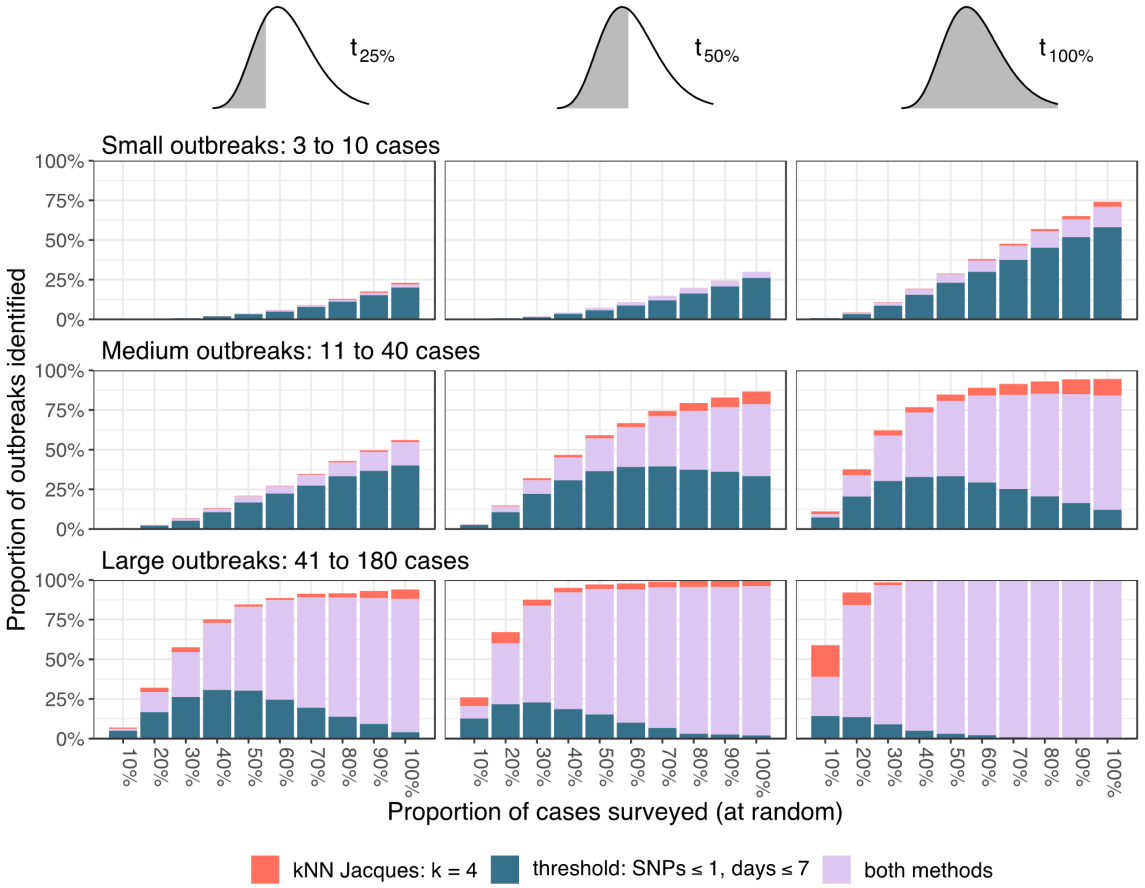

Figure 2: Proportion of simulated outbreaks that were resolved averaged over many repeats of the simulation and sampling steps. The columns show the resolution at different time points in the outbreaks, with $t_{25\%}$ meaning the day by which a quarter of the outbreak's cases had been reported. Rows show the size of outbreaks measured by total number of cases. The colours indicate which clustering methods contributed to the result; the total height of bars is the proportion identified.

Beyond calibrating the simulation's parameters to known characteristics of the virus, the simulation requires the outbreak context to be specified. For this study, we illustrated the method using a scenario modelled after the initial weeks of the Delta variant wave in NSW. We generated nine simultaneous outbreaks of different sizes originating from distinct index cases, which we generated with small random differences in viral genomes and starting dates similar in magnitude to the differences observed between subclusters of the Delta wave.

Assessing clustering

Using the pairwise distance data, we reconstructed clusters of cases based on two algorithms that exploit the fact that while a pair of unrelated cases might coincidentally be close in either temporally or genetically, cases that are simultaneously close under both measures are more likely to be causally related (Figure 1C-D). Next we check these clusters against our ground truth outbreak membership: we designate good quality clusters comprised of mostly cases from a single outbreak as "informative", whereas mixed or very small clusters are uninformative. Our goal was not to perfectly delineate outbreaks; indeed, if this were reliably possible using the pairwise distance data then the clusters would likely be sufficiently distinct that even sparse sampling would capture the circulating diversity.

Instead, we defined our measure of success per distinct outbreak: an outbreak is considered to be identified so long as there is at least one informative cluster associated with it.

Applying the method

With this automated method of determining adequate clustering resolution, we could test different sampling strategies. To illustrate the concept, we chose strategies that sampled a fixed proportion of cases at random. We then repeated the simulation and the sampling steps many times to average over random fluctuations. The results in Figure 2 show that small outbreaks with ten or fewer cases were difficult to resolve even with a high surveillance depth (horizontal axis). Larger outbreaks with more than forty cases could still be reliably resolved at 50% surveillance coverage before reaching their peak size, with performance dropping off sharply as the surveillance depth was decreased below 30%. We also tested how the performance varies as the genetic and temporal similarity between the index cases is altered. We applied the same approach to a historical dataset of cases from the NSW Delta variant wave, using the observed clusters as a ground truth to estimate the delays in first detecting each cluster had the surveillance depth been lower than it was in reality.

The results of this study illustrate the framework but are not directly applicable to other contexts; our intention is that the method should be adapted by altering the details of the simulation to suit anticipated scenarios. Once the simulation is validated, the framework allows different sampling strategies to be tested. We see this approach as providing complementary evidence to existing methods supporting the design of genomic surveillance programs for scenarios where public health objectives depend on high resolution clustering.

References

1. Capon, A., Sheppeard, V., Gonzalez, N., Draper, J., Zhu, A., Browne, M., Sullivan, E., Mihajlovic, M., Rockett, R., and Ferson, M. (2021). Bondi and beyond. Lessons from three waves of COVID-19 from 2020. Public Health Res. Pract. 31. 10.17061/phrp3132112.

2. Rockett, R.J., Arnott, A., Lam, C., Sadsad, R., Timms, V., Gray, K.-A., Eden, J.-S., Chang, S., Gall, M., Draper, J., et al. (2020). Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat. Med. 26, 1398–1404. 10.1038/s41591-020-1000-7.

3. Arnott, A., Draper, J., Rockett, R.J., Lam, C., Sadsad, R., Gall, M., Martinez, E., Byun, R., Musto, J., Marais, B., et al. (2021). Documenting elimination of co-circulating COVID-19 clusters using genomics in New South Wales, Australia. BMC Res. Notes 14, 415. 10.1186/s13104-021-05827-x.

4. Katelaris, A.L., Wells, J., Clark, P., Norton, S., Rockett, R., Arnott, A., Sintchenko, V., Corbett, S., and Bag, S.K. (2021). Epidemiologic Evidence for Airborne Transmission of SARS-CoV-2 during Church Singing, Australia, 2020. Emerg. Infect. Dis. 27. 10.3201/eid2706.210465.

5. European Centre for Disease Prevention and Control (2021). Guidance for representative and targeted genomic SARS-CoV-2 monitoring (ECDC).

6. World Health Organization (2021). Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health.

The Broad Street Pump

CIDM-PH Senior Investigator Profile

KATHY TANNOUS

Dr W. Kathy Tannous is an Associate Professor in Economics and Finance in the School of Business at Western Sydney University. She is a co-lead in Resilient Construction and Infrastructure Program at the Urban Transformation Research Centre and is a Senior Investigator in the Centre for Infectious Disease and Microbiology – Public Health. She has a strong research program focused on using linked health administrative and surveys data for predictive modelling. She is currently supervising 8 PhD students in total with 7 on industry sourced scholarships. Dr Tannous currently has grants valuing over $6.5 million with agencies including Fire and Rescue NSW, The Butterfly Foundation, NSW Health, and National Health and Medical Research Council. She has over 100 publications that comprise combinations of academic journal articles, industry reports, book chapters and books. Her research is highly applied with strong impact on policy and practices at all levels.

A/Prof Kathy Tannous
Western Sydney University
k.tannous@westernsydney.edu.au

News & Events

The Sydney Summer School in Pathogen Genomics and Global Health was successfully held for the held for the 7th consecutive year, co-hosted by the Centre for Infectious Disease and Microbiology – Public Health (CIDM-PH) and Sydney Institute for Infectious Diseases (Sydney ID), University of Sydney. The Summer School was held at the University of Sydney Business School on 13th – 17th February 2023. The 5-day program included a mix of inspiring keynotes, informatics reviews, masterclasses, practical hands-on exercises and a laboratory visit to the Genomics Sequencing Facility at the Centre for Infectious Diseases and Microbiology, ICPMR, Westmead.

Sydney Summer School in Pathogen Genomics & Global Health 2023

CONTACT US

CIDM-PH

Centre for Infectious Diseases and Microbiology Public Health

Westmead Hospital, ICPMR Building,
Level 3, Cnr Hawkesbury & Darcy Roads,
Westmead NSW 2145

PHONE:

(612) 8890 9870

WEBSITE:

www.wslhd.health.nsw.gov.au/CIDM-PH

E-MAIL:

WSLHD-CIDM-PH@health.nsw.gov.au

SYDNEY ID

Sydney Institute for Infectious Diseases

Westmead Hospital, Block K, Level 5

Cnr Hawkesbury & Darcy Roads
Westmead NSW 2145

WEBSITE:

www.sydney.edu.au/infectious-diseases-institute

E-MAIL:

infectious.diseases@sydney.edu.au

UPCOMING EVENTS

SAVE THE DATE…

PROGRAM & REGISTRATION COMING SOON

CIDM-PH Annual Colloquium

24 November 2023

9.00am – 5:00pm (AEST)

WECC, Westmead Hospital, Westmead NSW

Event Enquiries:

WSLHD-CIDM-PH@health.nsw.gov.au

CIDM centre for infectious diseases & microbiology Public Health

THE UNIVERSITY OF SYDNEY